ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Computer Science 17 (2013) 759 – 769
Procedia Computer Science

Information Technology and Quantitative Management (ITQM2013)

Energy Time Series Data Analysis based on a Novel Integrated Data Characteristic Testing Approach

Ling Tang[a]*, Chenghao Wang[b], Shuai Wang[c]

[a] *School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China*
[b] *Risk Management Department, Agricultural Bank of China, Beijing 100006, China*
[c] *NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

Abstract

This paper attempts to propose an integrated data characteristic testing approach for energy time series data so as to analyze the energy dynamics, which serves as the foundation for the model selection problem. Based on thoroughly analyzing the main data characteristics of energy time series data together with their interrelationship, these data characteristics are divided into two main categories: nature and pattern characteristics to explore energy time series data from different perspectives. In nature determination, the energy time series data is analyzed in terms of nonstationarity, nonlinearity and complexity characteristics from a global perspective. In pattern measurements, the characteristics of cyclicity (and seasonality), mutability (or saltation) and randomicity (or noise pattern) signify the relative hidden patterns and the impacts on the original data, via a way of decomposition. For illustration purpose, hydropower consumptions in China and USA are analyzed and the main data characteristics are thoroughly explored by using the proposed integrated approach. Empirical results reveal that besides same characteristics of difference strationarity, nonlinearity and seasonality, the hydropower markets in China and USA are quite different: while China's hydropower market are comparatively simple but sensitive to emergencies, e.g., government support and technological progress, US' hydropower market is otherwise mature and efficient with the nature of high leveled complexity and the main pattern of randomicity. The results also confirm the proposed integrated approach an effective tool to test energy time series data in terms of data characteristics, paving the way for the further model formulation and forecasting.

Selection and peer-review under responsibility of the organizers of the 2013 International Conference on Information Technology and Quantitative Management

Keywords: Data characteristics, Energy analysis, Time series data, Hydropower consumption

* Corresponding author. Tel.:86- 010-51452675.
E-mail address: tangling@mail.buct.edu.cn.

Selection and peer-review under responsibility of the organizers of the 2013 International Conference on Information Technology and Quantitative Management
doi:10.1016/j.procs.2013.05.098

1. Introduction

Facing the dilemma between resource shortage and environment destruction, numerous researches have been initiated within the field of energy study. For example, confronted with the fast increase in energy demand caused by economy growth, energy security has become a quite important issue, and lots of researches tried to capture the main trend in energy development, involving the productions, consumptions and prices of various energy forms [1-3]. Additionally, due to the difference in resource production and demand, many developed and developing countries, such as OECD, European countries, Japan, China and India, are looking for stable energy supplies from others with rich resources, and the risk in international energy market should not be ignored[4,5]. Moreover, since fossil energies are still the main resources in the world's total energy consumption, environment protection has also fallen into another hot topic, especially for the main developing country of China[6].

Undoubtedly, analysis for energy time series data in terms of data characteristics serves as the foundation for further researches mentioned above, especially prediction studies. For example, due to nonstationarity in energy consumption in Iran, Azadeh et al., (2010) proposed an integrated fuzzy algorithm for regression [7]. Liu (2011) tested the nonlinearity characteristic for petroleum price and accordingly formulated nonlinear model to forecast the future tendency [8]. Similarly, based on the inherent complexity in nuclear energy market, Tang et al., (2012) built a hybrid decomposition learning paradigm to forecast Chinese nuclear energy consumption [1]. Because of seasonality characteristic, Wang et al., (2011) introduced seasonal decomposition to partition the original data of hydropower consumption series data, then predicted each extracted component individually and finally fused all the results into the ultimate prediction [9]. Therefore, data characteristics testing should not be ignored as the premise in energy analysis study.

There have existed abundant studies on data characteristic testing for energy time series data. The main data characteristics in energy data mainly include: stationarity (and nonstationarity) [10], linearity (and nonlinearity) [8], chaotic property[11], complexity [1], fractality [11], regularity (and irregularity or randomness) [13], cyclicity [3], seasonality [9], saltation (or mutability) [14]. For example, Chen and Lee (2007) proposed a new panel unit root testing procedure to test the stationarity of energy consumption per capita [10]. Liu (2011) employed R/S analysis, power spectrum and the largest Lyapunov exponent to discover the nonlinearity of petroleum price [8]. Zhang et al., (2008) analyzed international crude oil price using a novel decomposition approach with empirical mode decomposition algorithm to capture the inner cyclical patterns [3].

However, most of these researches focused only on one data characteristic, ignoring others hidden in energy time series data. On the other hand, energy time series data have been proved to consist of several coexisting data characteristics at the same time [3]. Thus, even though testing for a single data characteristic can work whether the energy time series data processing such given characteristic, this is done at the risk of ignoring other potentially important characteristics together their interrelationship, significantly reducing the effectiveness of related researches. Therefore, an integrated data characteristic test approach is quite imperative to capture the predominant characteristic in the energy dynamic.

What's more, in existing modeling for energy researches, taking prediction for example, testing for the given data characteristic of the observed series data was rarely performed, not to mention an exploration for the main data characteristics from a integrated perspective. However, an effective model should not only be compatible with each data characteristic, but also consider the main characteristics of the observed data. Therefore, an integrated data characteristics testing approach is needed, not only to analyze the dynamics from a integrated perspective but also helpful for further researches (such as tasks of forecasting, simulation and programming) to determining the most appropriate model based on data characteristics.

Under such background, the main contribution of this paper is to formulate an integrated data characteristics testing approach for energy time series data in order to analyze energy system from an integrated perspective, which serves as the foundation for further other researches (e.g., prediction). The remaining part of this paper is organized as follows: Section 2 describes the proposed integrated data characteristic testing approach in detail. For illustration, an empirical study with the time series of hydropower consumptions in China and USA as samples is performed via the proposed approach, as discussed in Section 3. Section 4 concludes this paper and discusses the future researches.

2. Methodology Formulation

2.1. Data Characteristics

Generally speaking, data characteristics of energy time series data mainly include: stationarity (and nonstationarity), linearity (and nonlinearity), complexity, chaotic property, fractality, regularity (and irregularity), cyclicity, seasonality, saltation (or mutability), randomness and similar others.

Amongst various data characteristics of time series, stationarity (or nonstationarity) is the most basic data characteristic, especially for energy data [7,10]. Generally, this couple of characteristics is closely referred to the data generation process, where the main statistical properties of energy dynamics remain the same over time. More specifically, in a stationary time series data, any sub-data $x_t(t=n, n+1,\ldots,n+l)$might have a same distribution as its m-lag series data $x_t(t=n+m, n+m+1,\ldots,n+m+l)$, where n and l are arbitrary positive integer [15].

Besides, the dynamics of time series data can be generally divided into two main elementary categories: linear and nonlinear systems [16]. Due to interaction effects of numerous driving factors, energy dynamics can be considered as the latter one rather than the former, demonstrating nonlinearity characteristic. That is that energy time series data are driven by such systems for which it is difficult to govern the dynamics using simple linear technologies, where various driven factors of market, policy, technology, infrastructure and even emergency interact in a nonlinear manner [1].

Furthermore, recent researches have payed increasing attentions to diverse nonlinear characteristics, such as chaos, complex and fractality, within field of energy data study. Especially, chaos is related to the behavior of energy dynamic that seems superficially irregular but obeys certain hidden rules, sensitive only to the initial states [11]. Actually, a chaotic process is an intermediate state between regular and irregular state. The term complexity provides another description for chaos [17]. It is commonly accepted that a higher level of complexity indicates a more complicated dynamics with more irregular patterns [3]. Fractality, a geometry term, has also been introduced to describe the dimensional complexity of energy time series data in terms of various fractal dimensions, such as correlation dimension and information dimension [18].

It comes to our attention that although these diverse systems theories of chaotic, complex and fractal data characteristics aim at discovering individual data characteristics of energy time series, combined together they are much more useful in analyzing the complex state of energy dynamics, where regular (also termed determinate) and irregular (also termed random and stochastic) factors take places simultaneously. Furthermore, existing test technologies for these characteristics are quite similar and even blended with each other. Therefore, the data characteristics of chaotic property, complexity, fractality, regularity, irregularity and other similar ones fall into one category to explore the complexity state of energy system, which is termed complexity in this study. More specifically, a lower level of complexity in energy data time series is related to a strong power of the hidden rules governing the energy dynamics.

Usually, time series data, especially energy and economy data, comprise a series of coexisting hidden components, i.e., trend, cyclical, seasonal, saltatory and noisy patterns [3,19]:

$$x_t = f(t_t, c_t, s_t, m_t, n_t) \tag{1}$$

Where *tt*, *ct*, *st*, *mt* and *nt* denote the trendency, cyclical pattern, seasonal patten, saltatory pattern and noisy pattern hidden in time series data x_t at time *t*. Accordingly, the characteristics of cyclicity, seasonality, saltation (or mutability) and randomicity (or stochasticity) explore the respective impacts of their related hidden patterns on the observed time series.

Cyclical pattern, which returns to the beginning and repeats themselves in the same sequence with peaks and troughs, might be the most important component for energy time series data driven by certain hidden rules such as economic factors [20]. It is worth noticing that seasonal pattern can be seen as a special case of cyclical pattern with one year time scale, which is mainly caused by environmental factors (e.g. temperature, weather and climate)and socioeconomic ones (e.g. calendar and festival) [9]. Saltatory pattern generally is related to structural changes in energy time series data, mainly caused by some significant events [3, 14]. Besides, time series data is always corrupted by noise to different degree, and randomicity (or stochasticity) measures how much noise is there polluting the time series [13].

Based on the above discussion, the relationships across the main data characteristics of energy time series data can be captured, as illustrated in Fig. 1. The data characteristics of energy time series can be generally partitioned into two main kinds, i.e., the nature and pattern characteristics, which analyze energy time series data from different perspectives.

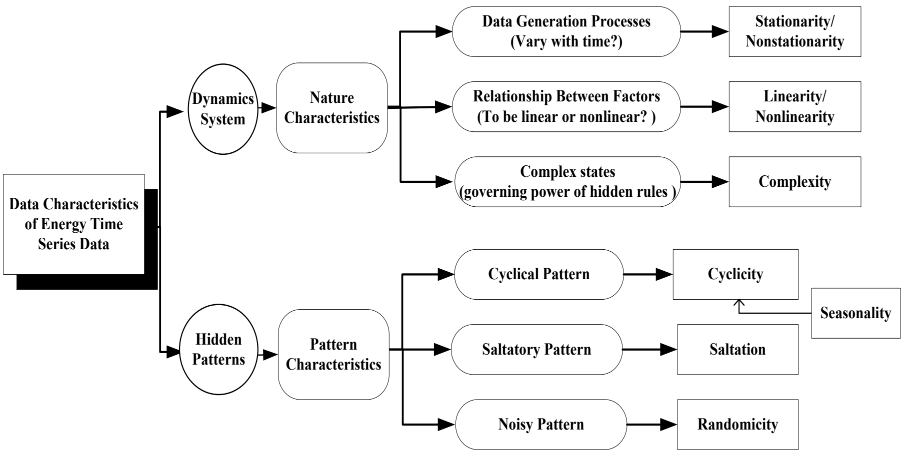

Fig. 1. Data characteristics of energy time series data.

It is worth noticing that though nature and pattern characteristics test energy time series data from distinct perspectives, they are actually closely related to and even dependent on each other. On one hand, nature characteristics view the energy time series data as a whole and determine inner structural rules in the dynamics system, while pattern characteristics focus on hidden components and analyze energy time series data via a way of decomposition. On the other hand, since the dynamics system consists of various hidden patterns, nature characteristics are actually dependent on the features of the extracted patterns together with the interrelationship with others, and similarly, patterns constitute the energy data in such a way governed by nature characteristics.

1) Nature characteristics

The characteristics of stationarity (and nonstationarity), linearity (and nonlinearity) and complexity (including fractality, chaos, irregularity and so on) are directly related to the energy dynamics, considering the energy time series data from a global perspective. Especially, stationarity and nonstationarity test the data generation processes driving energy time series data whether to remain or vary as time goes (without or with unit roots). Linearity and nonlinearity display relationships across various inner factors (and hidden patterns) within the energy systems. Complexity reflects the complex states of the energy dynamics systems in terms of the governing power of the deterministic rules (or regulation). Thus, these characteristics can be considered as nature characteristics, directly related to the essential property of the dynamics systems driving energy time series data.

2) Pattern characteristics

On the other hand, the data characteristics of cyclicity, seasonality, saltation and randomicity discover the relative hidden patterns and their respective impacts on the original data, exploring energy dynamics by decomposition approach. This approach assumes the energy data usually involves diverse hidden patterns at the same time. i.e., cyclical, seasonal, saltatory and noisy components; and these hidden patterns' impacts on the original time series are evaluated by the pattern characteristics of cyclicity, seasonality, saltation (or mutability) and randomicity (or stochasticity), respectively.

2.2. Main procedure of methodology

According to the analysis on the main data characteristics of energy time series data together with the relationships across them, an integrated data characteristic testing approach for energy time series data can be formulated. Accordingly, two main steps are involved in the proposed approach, i.e. nature determination and pattern measurement.

The proposed integrated testing approach first explores data from a global perspective in terms of nature characteristics, and further investigates hidden patterns from a more micro perspective. In nature characteristics, stationarity (or nonstationarity) is the most essential characteristic which should not be ignored in time series analysis, thus unit root test is performed firstly in the proposed scheme, and then nonlinearity and its excellent example complexity are considered. In pattern measurement, seeing that time series data possess series of patterns in the meantime, diverse pattern characteristics, i.e., cyclicity, saltation and randomicity, are estimated in a parallel rather than a successive manner. Therefore, with such structure, the main procedure of the proposed integrated approach can be summarized as follows.

Step 1: Nature determination

As the first step, the nature characteristics of the time series data are explored in terms of unit root tests, nonlinear tests and complexity exponent evaluation respectively, in order to analyze the whole energy dynamics. Especially, (1) stationarity is tested via unit root test to determine whether the data generation process driving energy dynamic stationary without unit root or nonstationary; (2) linearity via nonlinear test to determine whether the interrelationships across inner factors (or hidden patterns) in the energy dynamic systems follow a linear or nonlinear way; and (3) complexity in terms of exponent estimation to indicate how complicated the dynamics system is in terms of the governing power of determinative rules.

Step 2: Pattern measurement

Usually, energy time series data process a set of hidden patterns at the same time, i.e., cyclical (and seasonal) pattern, saltatory pattern and noisy pattern, and accordingly pattern characteristics of cyclicity (and seasonality), saltation and randomicity are estimated to not only check whether respective patterns exist in the observed data, but also measure the relative impacts on original series, in order to capture the main pattern characteristic in the energy time series data. Therefore, two sub-steps are involved in pattern measurement, i.e., pattern discovery and pattern importance estimation. Particularly, (1) in pattern discovery, relevant test and analysis are performed in order to confirm whether the given energy system covers the patterns of cyclical (including seasonal) pattern, saltatory pattern and noisy pattern; and then (2) through extracting or describing the related pattern, pattern importance is evaluated to display the respective impact on the original data and finally capture the main pattern characteristics, via quantitative forms, such as correlation coefficient and ratio of power, amplitude or density.

3. Empirical Study

3.1. Data Descriptions

Hydropower consumption time series data of China and USA are used as sample data, in this empirical study. The data are obtained from Wind Database (http://www.wind.com.cn/) and originally collected by China's National Bureau of Statistics and U.S. Energy Information Administration, respectively. The sample data are monthly data covering the period from January 1990 to June 2012 with a total of 270 observations, as shown in Fig. 2.

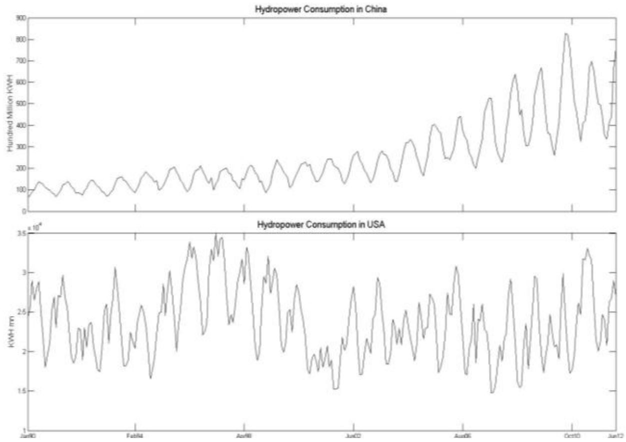

Fig. 2. The monthly hydropower consumption series time data in China and USA.

3.2. Empirical Results

3.2.1 Nature determination

For stationarity, the most popular unit root test, augmented Dickey-Fuller (ADF) test [21], is first utilized. In ADF test, data generation functions are determined in terms of Akaike information criterion (AIC) and the maximum of lag is set to 6 months. Additionally, KPSS test is then introduced as a complementary test to ADF test with the null hypothesis of stationarity [22]. Table 1 presents the stationary testing results for hydropower consumption of China and USA, marked as China_h and USA_h respectively.

Table 1. Stationarity characteristics of hydropower consumptions in China and USA.

Time Series	ADF Test				Stationarity
	Generation Functions (Constant, Trend, Lags)	*t-Stat*.	Critical Values (5% level)	*Prob*.	
China_h	(1,1,6)	-3.0030	-3.4270	0.1334	×
△(China_h)	(1,0,6)	-14.8164	-2.8724	0.0000	√
USA_h	(1,0,6)	-4.9029	-2.8723	0.0000	√
Time Series	KPSS Test				Nonstationarity
	Generation Functions (Constant, Trend, Bandwidth)	LM-Stat.	Critical Values (5% level)	Prob.	

China_h	(1,1,8)	0.5629	0.1460	<0.05	√
△(China_h)	(1,0,3)	0.0324	0.4630	>0.05	×
USA_h	(1,0,9)	0.3254	0.4630	>0.05	×

From the results, it can be seen that the sample data are of stationarity or difference-stationarity series. First, the existence of unit roots cannot be rejected in the hydropower consumption in China by ADF test, with the p-value is far above 5%. Furthermore, when measured by KPSS test, the null hypothesis of stationarity is rejected and the p-value is below 5%, otherwise accepting the alternative hypothesis that the hydropower consumption in China demonstrates strongly nonstationary at the confidence level of 95%. However, the tests prove that the one difference series of hydropower consumption in China demonstrate stationary.

Focusing on the hydropower consumption in USA, the results are totally different. First, KPSS test cannot reject the null hypothesis of stationarity of hydropower consumption in USA, since that the p-value exceeds 5%. Additionally, the ADF test further rejects the existence of unit roots and it indicates that the time series data of US' hydropower consumption is statonary with a confidence level of 95%.

For nonlinearity, surrogate data method is employed in this study. In the surrogate data based method, Fourier transform (FT) are employed, and the surrogating data are generated by multiplying the FT of the data by random phases and then transforming back to the time domain [23]. Then, the surrogate data and the original data are compared via Kolmogorov–Smirnoff (KS) test in terms of distributions [24]. According to above steps, a total of 1000 surrogates with different random phases are generated for nonlinearity tests. Table 2 reports the corresponding results, where *h*, *KS-Stat*. and *Prob*. present the ratio of the null hypothesis rejecting in all simulations at confidence levels of 95%, the mean calculated *KS* statistic and the mean p-value, respectively. From the table, it can be seen that the mean p-value are all far below 5%, indicating that statistically all the observed time series of hydropower consumptions in China and USA show nonlinearity at the 95% confidence level.

Table 2. Nonlinearity characteristics of hydropower consumptions in China and USA.

Time Series	**Surrogate Data Method**			**Linearity**
	h	*KS-Stat*	*Prob.*	
China_h	0.9980	0.2977	0.0003	×
US_h	0.9160	0.1616	0.0149	×

For complexity, complexity exponents are built based on the most popular fractal dimensions, e.g., correlation dimensions, via G-P algorithm [25]. In phase-pace reconstruction, the lag is determined via autocorrelation analysis, and embedding dimension m via the approach by [26]. Table 3 presents the corresponding results. It can be seen that while the complexity exponent of hydropower consumption in China is about 1.1520, the figure for USA is 2.7115 far above the former, implying that the hydropower consumption in USA might to be at an extremely higher complexity level compared with that of China.

Table 3. Complexity characteristics of hydropower consumptions in China and USA.

Time Series	Correlation Dimension Measurement			Complexity Level
	Lag (τ)	Embedding Dimension (m)	Correlation Dimension (D)	
China_h	6	3	1.1520	Low
US_h	6	9	2.7115	High

3.2.2 Pattern measurement

Cyclicity and seasonality are the most pattern characteristics in energy time series data. In this empirical studies, autocorrelation analysis is employed to capture the main time scale of the hidden cycle pattern, by showing how one point relates to the others in the interval of a given lag [27]. From the results, an obvious point can be found that a cyclical pattern with time windows of 12 months (one year) is hidden in both hydropower consumption series of China and USA, which can also be treated as seasonality. Especially, the absolute values of autocorrelation coefficients with lag of 12 months are far higher than others in all sample data, with the exceptions of one month lag in the original data which can be seen as trend other than cyclicity.

As far as saltation, ICSS test, in the CUSUM test families, is first employed for the residuals of linear ARMA fits built base on AIC. The point with the largest ICSS statistic is selected as the candidate point. Then, the original data is segmented into two subsets before and after the possible break point. Moreover, Chow tests are also employed to test the similarity between the two subsets [28]. If statistical test show that the point is the break point at 95% confidence level, the two subset series are tested for structural change in a similar way respectively, and these steps repeat until no break point can be identified at 95% confidence level. Table 4 reports the results for saltation tests, which indicate that the hydropower consumption series of China demonstrates obvious saltation, since the point of April in 2006 are tested as one structural break at 95% confidence levels. On the other hand, the results show that there might not any obvious break to change the data of hydropower consumption in USA structurally.

Table 4. Mutability characteristics of hydropower consumptions in China and USA.

Time Series	Point		ICSS	Chow	Log likelihood ratio	Wald	Structural Change
China_h	1996.12	*Stat.*	2.4483	0.8323	5.1799	11.7578	×
		prob.	0.0072	0.5458	0.5458	0.0676	
	2006.04	*Stat.*	5.6213	5.7589	33.9251	59.2500	√
		prob.	0.0000	0.0000	0.0000	0.0000	

To capture main pattern characteristics, dummy variables are employed for extracting the relative patterns. More specifically, for cyclical pattern, seasonal autoregressive (i.e. DC_t) model is introduced [29]. To model saltation, variables DU_t and DT_t presenting the structural change in the level and slope occurring at t_b (i.e. DU_t=1 if $t>t_b$, 0 otherwise; $DT_t=t-t_b$ if $t>t_b$, 0 otherwise) are added [30]. Therefore, the series data of hydropower consumption in China and USA can be described as:

$$x_t = f(t, DC_t, DU_t, DT_t) + \varepsilon_t \tag{2}$$

where ε_t is the series of residual errors, which can be treated as the noisy pattern. For modeling the observed data in terms of function (12), least square support vector regression (LSSVR), a powerful artificial intelligence (AI) technique, has been proposed [1]. Table 5 reports the analysis results accordingly.

Table 5. Main pattern characteristics of hydropower consumptions in China and USA.

Time Series	Statistic	Pattern Importance			Main Pattern Characteristic
		Cyclicality	Mutability	Randomicity	
China_h	Variance Ratio	13.90%	38.66%	6.82%	Saltation
	Correlation Coefficient	0.3791	0.6048	0.3523	
USA_h	Variance Ratio	46.31%	-	48.55%	Randomicity/ Cyclicality
	Correlation Coefficient	0.6997	-	0.7038	

From the results, it can be seen that the main pattern characteristic of hydropower consumption in China is saltation, followed by cyclicality, while randomicity and cyclicality are the main pattern characteristics for hydropower consumption in USA. Focusing on hydropower consumption in China, the variance ratio of saltatiory pattern to original data is about 38.66% and the correlation coefficient with the original data is the highest about 0.6048. As for hydropower consumption in USA, the main patterns are noisy and cyclical patterns, since their variance ratios are alike about 48.55% and 46.31%, and correlation coefficients about 0.7038 and 0.6997, respectively.

3.2.3 Discussion

Via the proposed integrated testing approach, the main data characteristics of hydropower consumption in China and USA are comprehensively analyzed, as illustrated in Table 6.

Table 6. Main data characteristics of hydropower consumptions in China and USA.

Time Series	**Nature Characteristics**			**Pattern Characteristics**		
	Stationarity (or Difference-Stationarity)	Linearity	Complexity	Cyclicity	Mutability	Randomicity
China_h	√	×	Lower		√	
USA_h	√	×	Higher	√		√

For nature characteristics, it can be seen from the testing results that the dynamics of hydropower consumption series in both China and USA demonstrate basic characteristics of stationarity (or difference-stationary) and nonlinearity. However, the complexity states in these two series are distinct, since that complex exponent of hydropower consumption in China approximates 1 while that of USA is otherwise far above 2. As far as pattern characteristics, hydropower consumption in China demonstrates strong mutability, indicating that the data are inclined to be changed by certain sudden changes. On the contrary, hydropower consumption in USA holds randomicity as the main pattern characteristic, closed followed by cyclicity.

Some even more interesting and valuable conclusions can be deduced from the testing results. First, besides stationarity and nonlinearity, the nature characteristic of complexity implies that the hydropower markets in China and USA are different from each other. Especially, high leveled complexity of US' hydropower consumption is associated with the efficiency of hydropower market in USA. This implies that the hydropower market structure of USA is complicated with diverse competing groups, and hydropower energy can be seen as one form of commodity driven by the various market factors, e.g., economic development, population growth, price of hydropower energy vis-a-vis other form. On the other hand, according to the low level of complexity, China's hydropower market otherwise is much less developed and the competitive level in the market is relatively lower, which is mainly driven by policy factors.

From the results of pattern characteristics, saltation further indicates that China's hydropower is extremely sensitive to policy factors, e.g., government support and technological progress. Especially, due to the ambitious hydropower development program of "Tenth Five-Year Plan" (2001-2005) and the "Eleventh Five-Year Plan" (2006-2010) and the outset of Three Gorges Hydropower Station, China's hydropower is taking a rapid development and the consumption has been largely enhanced around 2006, where an obvious break took place and changed the series data structurally. On the other hand, randomicity hidden in US' hydropower consumption further confirm the efficiency in hydropower market of USA, where lots of uncertain factors interact under such a complex market structure.

4. Conclusion and Future Research

This paper proposed an integrated data characteristics testing approach for energy time series data, in order to analyze the energy dynamics from a comprehensive perspective. Through discussing the main data characteristics of energy time series data carefully, the relationships across different data characteristics are analyzed. The main data characteristics are divided into two main categories: nature and pattern characteristics, which explore energy time series data from different perspectives.

Accordingly, the main steps are involved in the proposed approach, i.e., nature determination and pattern measurement. First, the nature characteristics of stationarity (and nonstationarity), linearity (and nonlinearity), complexity (including fractality, chaotic, irregularity and so on) are directly related to the energy dynamics, analyzing the energy time series data from a whole perspective. On the other hand, the pattern characteristics of cyclicity, seasonality, saltation (mutability) and randomicity discover the relative hidden patterns and their impacts on the original data, exploring energy time series data by an approach of decomposition.

For illustration purpose, an empirical study is performed with hydropower consumptions in China and USA as sample data, and the empirical results indicate that the proposed integrated scheme can be used as an effective tool to explore the main data characteristics of energy time series data. In nature determination, the complexity states in these two series are distinct, implying that the hydropower markets in China and USA are quite different. The low level of complexity indicates that China's hydropower market might still remains immature and lacks competitiveness, mainly driven by policy factors rather than market factors. On the other hand, US' hydropower market is otherwise mature and efficient with the nature of high leveled complexity. For pattern characteristics, saltation in China's hydropower consumption further confirms that the hydropower market in China is extremely sensitive to policy factors, such as government support and technological progress. On the other hand, randomicity, as the main pattern characteristic of US' hydropower consumption, again stresses the efficiency in hydropower market of USA, where lots of uncertain factors interact under a complex market structure.

It is worth noticing that the integrated data characteristic testing approach serves as the foundation for further model formulation of the observed data for further researches, especially prediction. An effective model should not only conform to each main inner rules (mainly governed by nature characteristics), but also incorporate the main characteristics of the observed energy data (mainly displayed in terms of pattern characteristics). The work done in this paper paves the way for the development of such model, which is the future work as the next step naturally.

Acknowledgements

This work is supported by grants from the National Science Fund for Distinguished Young Scholars (NSFC No. 71025005) and the National Natural Science Foundation of China (NSFC No. 71201054, NSFC No. 90924024 and NSFC No. 91224001).

References

[1] Tang, L., Yu, L. A., Wang, S., Li, J. P., Wang, S. Y.A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting. *Applied Energy*. 2012; **93**, 432-443.

[2] Xie Q., Bo,X., Peng,S., Li,J.P., Xu.W., Han,H. Bandwidth Empirical Mode Decomposition and its Application.. International Journal of Wavelets, *Multiresolution and Information Processing*. 2008,**6(6)**:777-798.

[3] Zhang, X., Lai, K. K., Wang, S.-Y. A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Economics*. 2008; *30*, 905-918.

[4] Li, J. P., Tang, L., Sun, X. L., Yu, L. A., He, W., Yang, Y. Y. Country risk forecasting for major oil exporting countries: A decomposition hybrid approach. *Computers & Industrial Engineering*. 2012; **63**, 641-651.

[5] He, K., Xie, C., Chen, S., Lai, K. K. Estimating VaR in crude oil market: A novel multi-scale non-linear ensemble approach incorporating wavelet analysis and neural network. *Neurocomputing*.2009;**72**, 3428-3438.

[6] Bao, Q., Tang, L., Zhang, Z., Wang, S. Impacts of border carbon adjustments on China's sectoral emissions: Simulations with a dynamic computable general equilibrium model. *China Economic Review*. 2013; **24**, 77-94.

[7] Azadeh, A., Saberi, M., Seraj, O. An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: A case study of Iran. *Energy*. 2010; **35**, 2351-2366.

[8] Liu, L. Nonlinear Test and Forecasting of Petroleum Futures Prices Time Series. *Energy Procedia*. 2011; **5**, 754-758.

[9] Wang, S., Yu, L., Tang, L., Wang, S. A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China. *Energy*.2011; **36**, 6542-6554.

[10] Chen P. F., Lee C. C. Is energy consumption per capita broken stationary? New evidence from regional-based panels.*Energy Policy*.2007;**35(6)**, 3526-3540.

[11] Wang J. Z, Chi D. Z, Wu J, Lu V. W. Chaotic time series method combined with particle swarm optimization and trend adjustment for electricity demand forecasting. *Expert Systems with Applications*. 2011; **38**, 7(6): 8419-8429.

[12] Chang, X., Liu, X., Zhou, W. Hydropower in China at present and its further development. *Energy*.2010;**35**, 4400-4406.

[13] Hu, J., Gao, J. B., White, K. D. Estimating measurement noise in a time series by exploiting nonstationarity. *Chaos, Solitons & Fractals*. 2004; **22**, 807-819.

[14] Lee C. C., Chang C. P. Structural breaks, energy consumption, and economic growth revisited: Evidence from Taiwan. *Energy Economics*. 2005;**27**, 6 (11), 857-872.

[15] Nelson C. R., Plosser C. I., Trends and random walks in macroeconomics time series: some evidence and implications, *Journal of Monetary Economics*.1982; **10**, 139-162.

[16] Pena D., Rodriguez J., Detecting nonlinearity in time series by model selection criteria, *International Journal of Forecasting*.2005; **21** , 731– 748.

[17] Memon Q. A., Synchronized choas for network security, *Computer Communications*.2003; **26**, 498–505

[18] Chang T. P., Ko H. H., Liu F. J., Chen P. H., Chang Y. P., Liang Y. H., Jang H. Y., Lin T. C., Chen Y. H. Fractal dimension of wind speed time series. *Applied Energy*. 2012;**93**(5), 742-749.

[19] Silva A. D., Hyndman R. J., Snyder R., A multivariate innovations state space Beveridge-Nelson decomposition, *Economic Modeling*.2009; 2**6**,1067-1074.

[20] Zarnowitz V., Ozyildirim A., Time series decomposition and measurement of business cycles, trends and growth cycles, *Journal of Monetary Economics*.2006; **53**,1717-1739.

[21] Dickey D. A., Fuller W. A., Likelihood ratio statistics for autoregressive time series with a unitroot, *Econometrica*.1981; **49**,1057-1072

[22] Kwiatkowski D., Phillips P. C. B., Schmidt P., Shin Y., Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*. 1992; **54**, 159-178

[23] Theiler J, Longlin A, Galdrikian B, Farmer D. D., Testing for nonlinearity in time series: the method of surrogate data, *Phys. D*.1992;**58** ,77-94.

[24] Unsworth C. P., Cowper M. R., McLaughlin S., Mulgrew B., A new method to detect nonlinearity in a time-series: synthesizing surrogate data using a Kolmogorov–Smirnoff tested, hidden Markov model, *Phys. D* .2001;**155**,51-68.

[25] Grassberger P., Procaccia I, Measuring the strangeness of strange attractors, *Physica D*.1983; **9**,189-208.

[26] Kennel M. B., Brown R, Abarbanel H. D. I., Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physical Review*.1992; **45**(6), 3403-3411.

[27] Hart T., Coulson T., Trathan P. N., Time series analysis of biologging data: autocorrelation reveals periodicity of diving behaviour in macaroni penguins, *Animal Behaviour*.2010;**79** (4),845-855.

[28] Hušková M., Kirch C., PraškováZ.,Steinebach J., On the detection of changes in autoregressive time series II. Resampling procedures, *Journal of Statistical Planning and Inference*.2008; **138**,1697-1721.

[29] Yang L., Tschernig R.Non- and semiparametric identification of seasonal nonlinear autoregression models. *Econometric Theory* 2002;**18**, 1408-1448.

[30] Zivot E., Andrews D., Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis, *Journal of Business and Economic Statistics*. 1992; **10**,251-270.